United States Patent [19]

Brayton

[11] **4,177,448**

[45] **Dec. 4, 1979**

[54] **CHARACTER RECOGNITION SYSTEM AND METHOD MULTI-BIT CURVE VECTOR PROCESSING**

[75] Inventor: **Robert K. Brayton,** Briarcliff Manor, N.Y.

[73] Assignee: **International Business Machines Corporation,** Armonk, N.Y.

[21] Appl. No.: **919,535**

[22] Filed: **Jun. 26, 1978**

[51] **Int. Cl.**[2] **G06K 9/00;** G06K 9/16

[52] **U.S. Cl.** ..... **340/146.3 AE;** 340/146.3 D

[58] **Field of Search** ..... 340/146.3 AE, 146.3 AC, 340/146.3 AQ, 146.3 D, 146.3 Q, 146.3 T

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,601,803	8/1971	Ullmann	340/146.3 AQ
3,813,646	5/1974	Quarmby	340/146.3 AE
3,979,722	9/1976	Sakoe	340/146.3 AE
4,014,000	3/1977	Uno et al.	340/146.3 Q

Primary Examiner—Joseph M. Thesz

Attorney, Agent, or Firm—John J. Goodwin

[57] **ABSTRACT**

An automatic character recognition system and method for identifying an unknown character which is one of a class of known characters. The system is set up using known specimen characters from a large character training set which must first be selected based on the use to which the character recognition will be put. Using the selected set, features and shapes of the character vocabulary in the set are obtained using selected feature scan parameters, are processed as a plurality of representative and normalized pieces of curves and are then stored in the form of binary coded representations. The system set-up also includes selecting and storing canonic shape parameters. The canonic shapes are separate pieces or segments of lines and curves which are selected on the basis that their shapes can be found as component parts or within a significant number of the characters within the character set. Having selected a character training set, selected and stored feature scan parameters and selected and stored canonic shape parameters, the system set-up procedure is completed. The next procedure is referred to as "system training" in which the individual characters within the large character training set are processed with the prior knowledge of the identity of each character being processed. This consists of an individual curve following for each of the plurality of characters on the character set and recording the path coordinates resulting from the curve following operation. The path coordinates are matched against the stored canonic shape parameters and the "best match" features are encoded. Statistical tables are then formed based on the best match relationships between the known training characters and the canonic shapes. The character recognition system is now capable of hereinafter operating with and identifying unknown characters belonging to the recognition space character set. In this procedure, the unknown character is examined using feature scan parameters to extract the features from the unknown character, providing complex vectors of the measured path coordinates of the extracted features which are matched against the stored canonic shape parameters by computing complex inner products, and a best match feature is determined.

Finally, a plurality of row vectors are extracted from the statistical tables and combined to form a product vector. The largest component of the product vector is selected, and the column index j of the maximum component is noted. The unknown character is then identified as being a member of the character membership class whose column index is j.

10 Claims, 12 Drawing Figures